PERGAMON

Vision Research 40 (2000) 1639–1643

Vision Research

www.elsevier.com/locate/visres

Rapid communication

Color, form and luminance capture attention in visual search

Massimo Turatto *, Giovanni Galfano

Dipartimento di Psicologia dello Sviluppo, Università di Padova, Via Venezia 8, 35131 Padova, Italy

Received 9 November 1999; received in revised form 23 February 2000

Abstract

Extant models of visual attention predict that a salient element should produce a bottom-up activation leading to a stimulus-driven attentional capture (e.g. Cave, 1999). However, apart from onset, previous works manipulating set-size in visual search failed to provide empirical evidence for this kind of capture. By varying target-singelton distance method, based on a single set-size, we explored whether, in a serial search task, an attentional capture is triggered by static discontinuities such as those generated through the manipulation of color, form, and luminance. The results suggest that those physical properties are indeed able to capture attention automatically.

Keywords: Attentional capture; Visual search; Set-size method

1. Introduction

Following notions coming from recent models of visual selection (e.g. Cave & Wolfe, 1990; Wolfe, 1994, 1996; Cave, 1999), attention will be deployed in the visual field on the basis of the observer's goals and beliefs. Central to the present work is the notion that this should happen also accordingly to stimulus saliency (Koch & Ullman, 1985). Several studies were devoted to the investigation of such mechanisms. Although the goal-directed component has been well documented (e.g. Folk, Remington & Johnston, 1992), very surprisingly, many failed to report evidence for a real stimulus-driven attentional capture for static discontinuities (e.g. Jonides & Yantis, 1988; Folk & Annett, 1994; Todd & Kramer, 1994). Basically, what emerges in the literature is that only the dynamic discontinuities, such as abrupt onsets, grab attention automatically (Yantis, 1996, but see Martin-Emerson & Kramer, 1997), whereas color, form or luminance do not (Jonides & Yantis, 1988). The present study was aimed at trying to conciliate experimental data from the literature with theoretical models which predict stimulus-driven attentional capture even by an irrelevant static singleton (e.g. Cave, 1999). From our standpoint, it is possible that the lack of evidence for this phenomenon is due to the procedure often used to investigate this issue, that is to say, the display-size method. We decided to address this issue by means of a different method relying on a single set-size and monitoring target-singleton distance.

The task consisted in detecting the presence or absence of a vertical target line among differently oriented distractor lines. In the first experiment we manipulated color as a feature potentially able to produce a bottom-up attentional capture. In the second and third experiments we manipulated form and luminance, respectively.

2. Methods

Since we used very short exposure times, we chose accuracy as the only dependent variable, as suggested by Wolfe (1998). Also, accuracy was preferred because unexpected events can influence decision processes, which in turn can affect response latencies (see Gibson & Jiang, 1998).

In experiment 1 (see Fig. 1 panel A) stimuli consisted of six disk elements (1.6° in diameter), equally spaced

* Corresponding author.
E-mail address: mturatto@psico.unipd.it (M. Turatto)

0042-6989/00/$ - see front matter
PII: S0042-6989(00)00061-4

around a fixation point on an imaginary circle with a radius of 4.5°. For each display element, there was one red singleton disk among five green disks. The disk elements appeared on a dark background (0.15 cd/m^2).

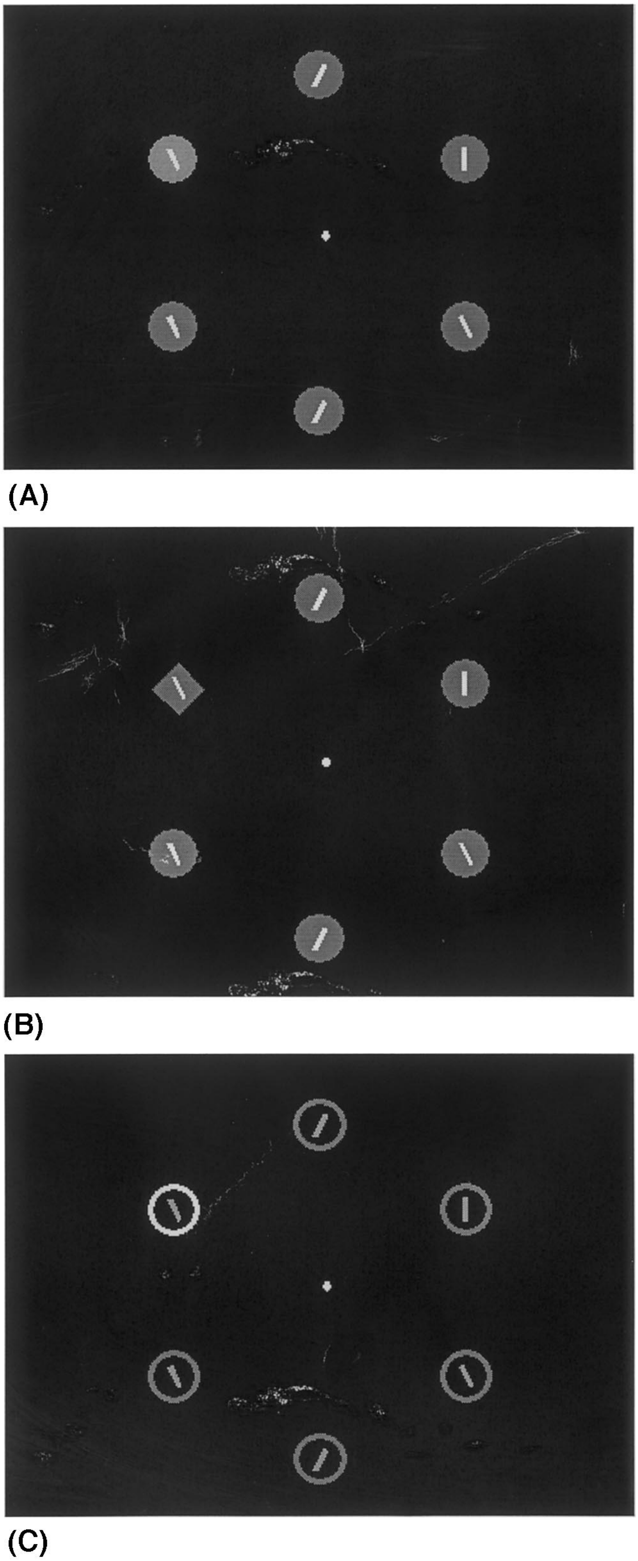

Fig. 1.

The target consisted in a vertical line segment embedded among distractors composed of randomly tilted lines, each oriented either to the left or to the right side. Target and distractor lines appeared inside the disk elements and were white lines (25 cd/m^2) covering 1.2° of visual angle. The degree of orientation of the distractors was adjusted according to an adaptive staircase procedure based on subjects' performance. This procedure was adopted in order to keep overall performance at 75% correct. This procedure also ensured that the target did not pop out, and the task was performed reasonably serially[1].

Target position in the array was random such that, when present, the target appeared inside the color singleton in only 1/6 of the total trials. As mentioned before, instead of varying set-size, we used the distance method, which basically consists in monitoring target position with respect to the singleton location (for a similar paradigm, see Cave & Zimmerman, 1997). We labeled position 0 (p0) the condition in which the target was inside the color singleton, p1 the condition in which the target flanked the singleton (to either the right or the left side), and p2 flanked that. Position p3 meant that the target was at the opposite side of the display.

Each subject performed one session composed of four blocks of 60 trials each. Before the beginning of

[1] We conducted a pilot experiment in which the dependent variable was represented by reaction times (RTs). The only difference with experiment 1 was that no singleton was presented. Ten students performed a visual search task reporting the presence or absence of the vertical target line. RTs were entered into a one-way repeated measure ANOVA, with target presence (target present or absent, regardless of position) as the only factor, which resulted in a significant effect, $F(1,9) = 27.804$, $P < 0.005$. By dividing the RTs difference (107 ms) between target present and target absent conditions by 1/2 of the number of elements (six) in the array, we estimated that the time required for processing each item was about 35 ms/item, a typical value which is consistent with that of those tasks in which the target does not pop-out and the search is not efficient (Wolfe, 1998). This allowed us to reasonably assume that subjects did not perform the task by means of a singleton detection mode (Bacon & Egeth, 1994). In addition, we used two kinds of distractors, a precaution which presumably compelled subjects to adopt a feature search mode strategy (Wolfe, Friedman-Hill, Stewart & O'Connell, 1992).

Fig. 1. Panels A, B and C illustrate the display stimuli used in color, form and luminance experiments, respectively. In each condition there was a salient element that differed from the others on the basis of a specific dimension. For the color condition (panel A) differences in color are represented as differences in gray level (bright gray corresponds to red, dark gray corresponds to green). However, in all the experiments target position (vertical line) was randomly correlated with that of the singleton, that is to say that target appeared inside the singleton only on 1/6 of the total trials. In all the examples target is presented inside the element at 120° (p2) from the singleton. Non-target elements (tilted lines) varied their orientation according to subjects' performance in order to render the task a serial visual search.

the experiment, subjects performed a training session, consisting in a number of trials that were carried out until subjects felt confident with the task. Trials began with a fixation point presented for 500 ms, then the visual search display was turned on and displayed only for 180 ms, rendering any eye movements useless. From stimulus onset, subjects had 2500 ms for responding. Half of the subjects responded to target present with the left hand ('Q' key), and to target absent with the right hand ('P' key); the remainders vice-versa. The feedback for the incorrect responses was a 500 ms, 500 Hz tone, presented together with the message 'error'. If a response was not produced within 2500 ms, the same sound signal, along with the display message 'missed response', was presented. Subjects were told to be as accurate as possible in making their responses.

In experiment 2, the task was identical to that used in experiment 1, but the stimuli consisted in one red diamond (1.6°) among five red disks (see Fig. 1 panel B).

In experiment 3, the task was identical to that used in experiments 1 and 2 except that the stimuli consisted in six gray outline disks presented on a dark background. The line segments (including the target) appeared in red. In each display, there were one bright-gray disk among five dim-gray disks[2] (see Fig. 1 panel C).

Each experiment was performed by a different sample of subjects. Thirteen, ten and 11 subjects participated as naïve volunteers in experiments 1, 2 and 3, respectively.

All subjects were students at the University of Padova and had normal or corrected-to-normal vision. Subjects were not aware of the purpose of the experiment, and had not taken part in previous experiments on visual search.

3. Results

In all the experiments, accuracy data (proportions correct) for the Target present condition only were entered into an one-way repeated measures ANOVA, in which target position (four levels: p0, p1, p2, and p3) was the only factor considered.

[2] The photometric and colorimetric measurements were carried out by means of a Minolta chromameter CS-100. The green (CIE x, y chromaticity coordinates of 0.270/0.618; RGB palette value set at 0, 28, 0) and red disks (CIE x, y chromaticity coordinates of 0.598/0.347; RGB palette value set at 41, 0, 0) had a luminance of 2.0 cd/m^2. The bright-gray (CIE x, y chromaticity coordinates of 0.330/0.330; RGB palette value set at 60, 60, 60) element had a luminance of 20.0 cd/m^2, whereas the dim-gray (CIE x, y chromaticity coordinates of 0.330/0.330; RGB palette value set at 25, 25, 25) had a luminance of 2.0 cd/m^2.

3.1. Experiment 1: color

Target position was significant $F(3,36) = 18.276$, $P < 0.0001$. Pairwise comparisons (t-tests) revealed that proportion correct in p0 [$M = 0.794$, SD = 0.085] was significantly higher ($P < 0.001$) than in p1 [$M = 0.568$, SD = 0.109], p2 [$M = 0.541$, SD = 0.104], and p3 [$M = 0.565$, SD = 0.175]. Fig. 2 shows the percentage of target seen as a function of the four different levels of the target position factor in experiment 1, 2 and 3.

As Fig. 2 clearly suggests, a color singleton was able to draw attention automatically. Indeed, because color was completely task-irrelevant, it seems reasonable to assume that no attentional control setting for color was present in the task. This assumption is motivated on two grounds. First, the estimated processing time[1] for each item is not compatible with a pop-out process, which means that the vertical line target was not found using a singleton detection mode, that is a strategy for the odd-man out (Bacon & Egeth, 1994). Second, the color singleton was not useful for the task, in that it contained the target line only 1/6 of the total trials (Yantis, 1993). Therefore, the capture can be interpreted as purely stimulus-driven.

3.2. Experiment 2: form

Target position was significant $F(3,27) = 22.689$, $P < 0.0001$. Pairwise comparisons showed that proportion correct in p0 [$M = 0.785$, SD = 0.087] was significantly higher (all Ps < 0.001) than in p1 [$M = 0.549$, SD = 0.076], p2 [$M = 0.604$, SD = 0.085], and p3 [$M = 0.591$, SD = 0.079]. As in experiment 1, the results showed that performance when the target was inside the singleton was more accurate than in all other positions in the array. This finding provides further support for the hypothesis that form was able to capture attention in a

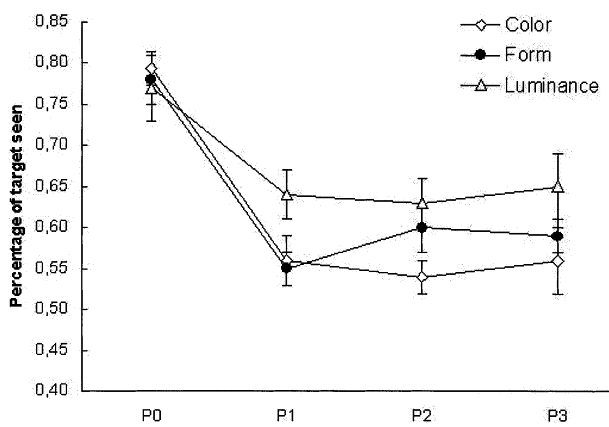

Fig. 2. Accuracy expressed as percentage of target seen in the three experiments. The patterns of results were very similar in experiments 1, 2 and 3, indicating that a salient element in the array defined either by color, form or luminance elicited an involuntary spatial shift of attention, which, in turn, increased subjects' performance for target detection. Bars represent ± 1 SE.

purely stimulus-driven manner. The present findings fly in the face of those of Theeuwes (1992, experiment 1A), which were taken as evidence that an element with a unique form did not attract attention (but see Theeuwes, 1994). His subjects searched for a target element differing from the other items in color. Under these conditions, the presence of a form singleton element did not slow down latencies for target detection. We argue that the discrepancy with our results could be due to the procedure employed in investigating the phenomenon. Theeuwes manipulated the number of elements displayed in the search array (display-size method). As we will point out in the discussion (see below), such procedure could be responsible for the 'masking' of the attentional capture phenomenon.

3.3. Experiment 3: luminance

Target position was significant $F(3,30) = 6.837$, $P < 0.001$. Pairwise comparisons revealed that proportion correct in p0 [$M = 0.775$, $SD = 0.146$] was significantly higher (all $Ps < 0.001$) than in p1 [$M = 0.640$, $SD = 0.093$], p2 [$M = 0.633$, $SD = 0.102$], and p3 [$M = 0.642$, $SD = 0.136$]. The results showed that a salient element defined in the luminance dimension attracted attention increasing subjects' performance for target detection. Again, this result is consistent with the prediction made by many models of visual attention, but it is incongruent with Jonides and Yantis' (1988) behavioral observations.

4. Discussion

A visually salient element in a given environmental context is conceived as being the object that is different from the others at least in one feature dimension, say form, color, luminance and so on. For instance, no one would doubt that a green element among red elements is a phenomenally salient object, and, in fact, often advertisement panels and other media like magazines or television try to attract our attention by means of colored words, phrases or objects. Also, for other species in the animal world, the concept of salience plays an important role, especially in the prey–predator relationship. Indeed, for different aims, both the prey and the predator usually try not to be a salient element in the respective visual field. To this purpose, nature has developed, for some species, a sophisticate mechanism that, when necessary, operates by minimizing the salience of an animal with respect to the background (the landscape), namely, 'camouflage'.

In the present study we have shown that focal attention can be automatically captured by a salient object in the color, form and luminance dimensions, thus providing evidence supporting many models of visual attention which predict such phenomenon (e.g. Wolfe, 1994, 1996; Cave, 1999). Experiment 1 showed that color is able to summon attention automatically, a result which is consistent with the findings of Nothdurft (1993), but is at odds with those of Jonides and Yantis (1988), Folk and Annett (1994) and Todd and Kramer (1994). In experiment 2, we demonstrated that form represents an attribute that can elicit a shift of focal attention, in contrast with what was found by Theeuwes (1992). Finally, from experiment 3, it emerged that also luminance can produce an involuntary attentional capture, contrary to what Jonides and Yantis (1988) have shown.

How can we reconcile our results with previous findings which failed to demonstrate attentional capture merely on the basis of salience? A viable explanation to account for the different results could refer to the different methods that have been employed for exploring the attentional capture issue. It could be speculated that perhaps the set-size method is not the more proficient way to investigate attentional capture by static discontinuities (i.e. variations within a dimension over space but not over time). In fact, in such procedure, the local singleton signal needs to compete with the transient signal produced by global onset of the whole display. In addition, it is worth noticing that this competition effect is strongly exacerbated as set-size increases, in that the number of competing transients swamping the singleton signal grows as set-size increases.

The signal coming from the distractors acts as noise that is added to the relevant signal due to peak of activation elicited by the feature singleton. It follows that the signal/noise ratio is not kept uniform throughout set-sizes. It should be observed that the transient produced by the onset of the whole display might overwhelm the signal coming from the feature singleton even in the present experiments. However, the use of a single set-size allowed us to keep the signal/noise ratio constant, rather than increasingly disadvantaging the feature singleton at larger set-sizes (also, see the Threshold Interrupt Hypothesis, Martin-Emerson & Kramer, 1997). Hence, what is suggested here is that varying the number of elements in the display could be less than ideal for exploring stimulus-driven attentional capture, in that the strength of the singleton is evaluated under conditions which are unbalanced for the strength of the global onset[3].

[3] We are not claiming that detection of a target line is easier with a small number of distractor tilted lines. Sagi and Julesz (1987) showed that a target line was detected more efficiently with a high number of distractor elements. It should be noted, however, that the experimental condition we used is not comparable with that employed by Sagi and Julesz, in that the arrangement of elements in our paradigm did not conform to the conditions in which a facilitator effect of background lines numerosity could be observed, that is to say, those created by implementing in the display high density and a short inter-element distance.

Acknowledgements

The authors wish to thank Nicola Puntin and Enea Pavone for their help in collecting the data. Thanks are also due to Adam Reeves, Roberto Dell'Acqua, and two anonymous reviewers for their helpful comments on a first draft of this manuscript, and to Sandro Bettella for technical support.

References

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception and Psychophysics*, *55*, 485–496.

Cave, K. R. (1999). The FeatureGate model of visual attention. *Psychological Research*, *62*, 182–194.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*, 225–271.

Cave, K. R., & Zimmerman, J. M. (1997). Flexibility in spatial attention before and after practice. *Psychological Science*, *8*, 399–403.

Folk, C. L., & Annett, S. (1994). Do locally defined feature discontinuities capture attention? *Perception and Psychophysics*, *56*, 277–287.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology*: *Human Perception and Performance*, *18*, 1030–1044.

Gibson, B. S., & Jiang, Y. (1998). Surprise! An unexpected color singleton does not capture attention in visual search. *Psychological Science*, *9*, 176–182.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception and Psychophysics*, *43*, 346–354.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.

Martin-Emerson, R., & Kramer, A. F. (1997). Offset transients modulate attentional capture by sudden onsets. *Perception and Psychophysics*, *59*, 739–751.

Nothdurft, H. C. (1993). Saliency effects across dimensions in visual search. *Vision Research*, *33*, 839–844.

Sagi, D., & Julesz, B. (1987). Short-range limitation on detection of feature differences. *Spatial Vision*, *2*, 39–49.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception and Psychophysics*, *51*, 599–606.

Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception*, *23*, 429–440.

Todd, S., & Kramer, A. F. (1994). Attentional misguidance in visual search. *Perception and Psychophysics*, *56*, 198–210.

Wolfe, J. M. (1994). Guided Search 2.0: a revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202–238.

Wolfe, J. M. (1996). Extending guided search: why guided search needs a preattentive 'item map'. In A. F. Kramer, M. G. H. Coles, & G. D. Logan, *Converging operations in the study of visual selective attention* (pp. 247–270). Washington, DC: American Psychological Association.

Wolfe, J. M. (1998). Visual search. In H. Pashler, *Attention* (pp. 13–73). Hove, UK: Psychology Press.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology*: *Human Perception and Performance*, *18*, 34–49.

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology*: *Human Perception and Performance*, *19*, 676–681.

Yantis, S. (1996). Attentional capture in vision. In A. F. Kramer, M. G. H. Coles, & G. D. Logan, *Converging operations in the study of visual selective attention* (pp. 45–76). Washington, DC: American Psychological Association.